(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 443 417 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
04.08.2004 Bulletin 2004/32

(51) Int Cl.7: **G06F 15/78**, G11C 16/30

(21) Application number: 03425054.8

(22) Date of filing: 31.01.2003

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PT SE SI SK TR
Designated Extension States:
AL LT LV MK RO

(71) Applicant: STMicroelectronics S.r.l.
20041 Agrate Brianza (Milano) (IT)

(72) Inventors:
- Borgatti, Michele
  41035 Finale Emilia (Modena) (IT)
- Cali, Lorenzo
  20045 Besana Brianza (MI) (IT)
- Lertora, Francesco
  16040 San Colombano Certenoli (Genova) (IT)
- Pasotti, Marco
  27028 San Martino Siccomario (Pavia) (IT)
- Rolandi, Pier Luigi
  15059 Monleale (Alessandria) (IT)

(74) Representative: Botti, Mario
Botti & Ferrari S.r.l.,
Via Locatelli, 5
20124 Milano (IT)

(54) A reconfigurable signal processor with embedded flash memory device

(57) The present invention relates to a dynamically reconfigurable processing unit (1) including an embedded Flash memory device (3) for non-volatile storage of code, data and bit-streams, the unit (1) being integrated into a single chip together with a microprocessor (2) core. Advantageously, the processing unit further comprises an S-RAM based embedded FPGA unit structured for FPGA reconfigurations having a specific programming interface (7) connected to a port (FP) of said Flash memory device (4) through a DMA channel (8).

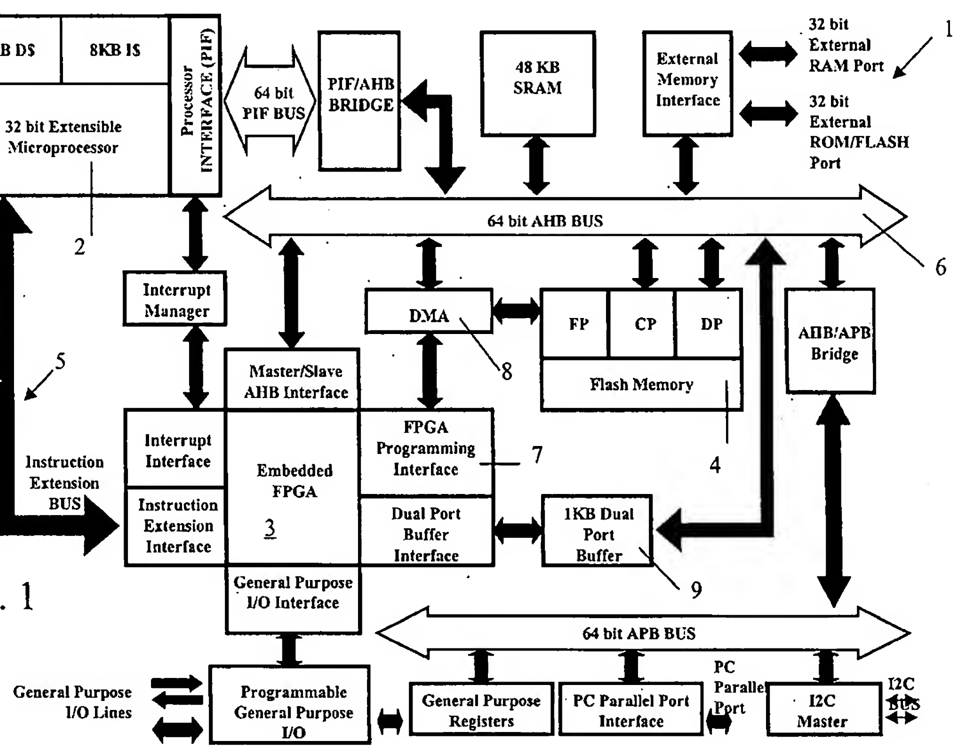


FIG. 1

EP 1 443 417 A1

Description

Field of invention

[0001] The present invention relates to a dynamically reconfigurable processing unit tightly connected to a Flash EEPROM memory subsystem.

[0002] More specifically, the invention relates to reconfigurable signal processing IC with an embedded Flash memory device for non-volatile storage of code, data and bit-streams, the unit being integrated into a single chip together with a microprocessor core.

Prior art

[0003] As is well know by those skilled in this technical field, increasing complexity of system design and shorter time-to-market requirements are leading research towards the investigation of hybrid systems including processors enhanced by programmable logic.

[0004] In this respect, reference is made to the work by Young-Don Bae et al., "A Single-Chip Programmable Platform Base on A Multithreaded Processor and Configurable Logic Clusters", ISSCC 2002 Digest of Technical Papers, pp 336-337, Feb. 2002.

[0005] Moreover, a further reference may be considered the article by Zhang et al., having title: "A 1V Heterogeneous Reconfigurable Processor IC for Baseband Wireless Applications", ISSCC 2000 Digest of Technical Papers, pp 68-69,488, Feb. 2000.

[0006] At the same time raising costs of mask sets and shorter time-to-market available for new products, are leading to the introduction of systems with a higher degree of programmability and configurability, such as system-on-chip with configurable processors, embedded FPGA and embedded flash memory.

[0007] Moreover, the availability of an advanced embedded flash technology, based on NOR architecture, together with innovative IP's, like embedded flash maccrocells with special features, is a key factor.

[0008] For a better understanding of the present invention reference is also made to the Field Programmable Gate Array (FPGA) technology combining standard processors with embedded FPGA devices.

[0009] These solutions allows to configure into the FPGA at deployment time exactly the required peripherals, exploiting temporal re-use by dynamically reconfiguring the instruction-set at run time based on the currently executed algorithm.

[0010] The existing models for designing FPGA/processor interaction can be grouped in two main categories:

- the FPGA is a co-processor communicating with the main processor through a system bus or a specific I/O channel;
- the FPGA is described as a function unit of the processor pipeline.

[0011] The first group includes the GARP processor, known from the article by T. Callahan, J. Hauser, and J. Wawrzynek having title: "The Garp architecture and C compiler" IEEE Computer, 33(4) : 62-69, April 2000. A similar architecture is provided by the A-EPIC processor that is disclosed in the article by S. Palem and S. Talla having title: "Adaptive explicit parallel instruction computing", Proceedings of the fourth Australasian COmputer Architecture Conference (ACOAC), January 2001.

[0012] In both cases the FPGA is addressed via dedicated instructions, moving data explicitly to and from the processor. Control hardware is kept to a minimum, since no interlocks are needed to avoid hazards, but a significant overhead in clock cycles is required to implement communication.

[0013] Only when the number of cycles per execution of the FPGA is relatively high, the communication overhead may be considered negligible.

[0014] In the commercial world, FPGA suppliers such as Altera Corporation offer digital architectures based on the US Patent No. 5,968,161 to T.J. Southgate, "FPGA based configurable CPU additionally including second programmable section for implementation of custom hardware support".

[0015] Other suppliers (Xilinx, Triscend) offer chips containing a processor embedded on the same silicon IC with embedded FPGA logic. See for instance the US Patent 6,467,009 to S.P. Winegarden et al., "Configurable Processor System Unit", assigned to Triscend Corporation.

[0016] However, those chips are generally loosely coupled by a high speed dedicated bus, performing as two separate execution units rather than being merged in a single architectural entity. In this manner the FPGA does not have direct access to the processor memory subsystem, which is one of the strengths of academic approaches outlined above.

[0017] In the second category (FPGA as a function unit) we find architectures commercially known as: "PRISC"; "Chimaera" and "ConCISe".

[0018] In all these models, data are read and written directly on the processor register file minimizing overhead due to communication. In most cases, to minimize control logic and hazard handling and to fit in the processor pipeline stages, the FPGA is limited to combinatorial logic only, thus severely limiting the performance boost that can be achieved.

[0019] These solutions represent a significant step toward a low-overhead interface between the two entities. Nevertheless, due to the granularity of FPGA operations and its hardware oriented structure, their approach is still very coarse-grained, reducing the possible resource usage parallelism and again including hardware issues not familiar nor friendly to software compilation tools and algorithm developers.

[0020] Thus, a relevant drawback in this approach is

often the memory data access bottleneck that often forces long stalls on the FPGA device in order to fetch on the shared registers enough data to justify its activation.

[0021] The technical problem of the present invention is that of providing a new kind of reconfigurable processing unit tightly connected to a memory architecture having functional and structural features capable to offer significant performance and energy consumption enhancements with respect to a traditional signal processing device.

Summary of invention

[0022] The invention overcomes the limitations of similar preceding architectures relying on an embedded device of different nature, and a new approach to processor/memory interface.

[0023] According to a first embodiment of the present invention, the reconfigurable processing unit targets image-voice processing and recognition application domains by joining a configurable and extensible processor core and an SRAM-based embedded FPGA.

[0024] More specifically, the processing unit according to the invention further includes an S-RAM based embedded FPGA unit structured for FPGA reconfigurations having a specific programming interface connected to a port FA of said Flash memory device through a DMA channel.

[0025] The features and advantages of the processing unit according to this invention will become apparent from the following description of a best mode for carrying out the invention given by way of non-limiting example with reference to the enclosed drawings.

Brief description of the drawings

[0026]

Figure 1 is a block diagram of a processing unit architecture for data processing according to the present invention;

Figure 2 is a block diagram of a Flash memory architecture embedded into the processing unit of Figure 1;

Figure 3 is a schematic view of system memory hierarchy provided by the present invention;

Figure 4 is a block diagram of a specific processor extension, for instance added DSP instructions examples;

Figure 5 is a block diagram of a further specific processor extension, for instance an optimized fixed-point calculation of the square root accounts;

Figure 6 is a table view showing the overall performance improvements for a face recognition task implemented by the processing unit of the present invention;

Figure 7 is a schematic chip micrograph.

Detailed description

[0027] With reference to the drawings views, generally shown at 1 is a processing unit realized according to the present invention for digital signal processing based on reconfigurable computing.

[0028] The processing unit 1 includes an embedded Flash memory device 4 for non-volatile storage of code, data and bit-streams and a further S-RAM based embedded FPGA unit 3 realized for the configuration purposes of the present invention.

[0029] More specifically, a 8Mb application-specific embedded flash memory device 4 is disclosed. The memory device 4 is integrated into a single chip together with a microprocessor 2 and the FPGA structure 3.

[0030] Advantageously, application-specific hardware units are added and dynamically modified by the embedded FPGA 3 reconfiguration. By implementing application-specific vector processing instructions the processing unit 1 shows a peak computing power of 1GOPS.

[0031] Efficient read-write-erase access to code, data and FPGA bitstreams is provided by the Flash memory device 4 based on a modular 8Mb, 4-bank Flash memory, as will be more clearly explained hereinafter.

[0032] The processing unit 1 comprises three content-specific I/O ports and delivers an aggregate peak read throughput of 1.2GB/s.

[0033] The system architecture 1 is illustrated in Figure 1.

[0034] The functional purposes of the embedded FPGA 3 are:

i) extension of the processor datapath supporting a set of additional special-purpose C-callable microprocessor instructions;

ii) bus-mapped coprocessors, connected to the system bus through a master/ slave interface;

iii) flexible I/O to connect external units or sensors with application-specific communication protocols.

[0035] Even though such different circuit purposes would require different kinds of programmable logic for best implementation of either arithmetic-dominated or control-dominated logic, a single programmable logic subsystem 3 has been implemented to be shared among different purposes both in space (same configuration) and time (subsequent configurations).

[0036] The single, high I / O count, fine-grain e-FPGA 3 operates as a datapath for the microprocessor pipeline

and as dedicated control logic for bus coprocessor and I/O control interface. The FPGA has a specific programming interface 7 connected to a port FP of said Flash memory device 4 through a DMA channel 8.

[0037] FPGA reconfiguration is concurrent to software execution.

[0038] A local bus 6 connects a dedicated 32-bit Flash memory port FP to the FPGA programming interface 7.

[0039] A DMA channel 8 handles the bitstream transfer while microprocessor fetches instructions and data from different Flash memory ports: 64-bit wide code port (CP) and data port (DP).

[0040] To support streaming applications a 1kB dual-port buffer 9 is used to interface fast decoding hardware and slower software running on the processor 2.

[0041] The memory sub-system architecture is shown in Figure 2.

[0042] The modular structure of the memory (dotted line) includes:

- charge pumps 10 (Power Block);
- testability circuits 11 (DFT);
- a power management arbiter 12 (PMA); and,
- a customizable array 13 of N independent 2Mb flash memory modules 16.

[0043] Depending on the storage requirements the number N may be chosen; N=4 in the current implementation.

[0044] The modular memory features (N+2) 128-bit target ports and implements a N-bank uniform memory 13.

[0045] As previously mentioned, three content-specific ports are dedicated to code (CP, 64-bit wide), data (DP, 64-bit) and FPGA bit stream configurations (FP, 32-bit). A 128 bit sub-system crossbar 15 connects all the architecture blocks and the eight bit microprocessor 2.

[0046] The main features of such the flash memory device 4 are: charge pump 10 sharing among different flash memory modules 16 through the PMA arbiter 12 in a multi-bank fashion. Moreover, the use of a small eight bit micro processor 2 to easy memory system test and to add complex functionalities for data management, and the use of an ADC (Analog-to-Digital Converter), required by the application, to increase system self test capability.

[0047] The third FP port of the Flash device 4 is dedicated to manage embedded-FPGA (e-FPGA) configurations data stored in flash memory modules. The FP port is read-only and provides fast sequential access for bit streams downloading.

[0048] The FP has four configuration registers replicating the information stored in CP port that must be used in order to write e-FPGA configurations data.

[0049] The output data word bus and the address bus are 32 bits wide. The FP port uses a chip select to access in the addressable memory space, and a burst enable to allow burst serial access.

[0050] In read operation, an output ready signal is tied low when data are not immediately available, so that it can acts as a wait state signal.

[0051] The eight-bit microprocessor 2 (uP) performs additional complex functions (defragmentation, compression, virtual erase, etc.) not natively supported by the DP port, and assists for built-in self test of the memory system.

[0052] The (N+2)x4 128-bit crossbar 15 connects the modular memory with the four initiators (CP, DP, FP and uP) providing that at least three flash memory modules 16 can be read in parallel at full speed.

[0053] The memory space of the four modules 16 is arranged in three programmable user-defined partitions, each one devoted to a port. The memory system clock can run up to 100MHz, and reading three modules 16 with 128bit data bus and 40ns access time, results in a peak read throughput of 1.2GB/s.

[0054] Each 2Mb flash memory module 16 has a 128-bit IO data bus with 40ns access time, resulting in 400Mbyte/s, and a program/erase control unit. Simultaneous memory operations use the power management arbiter 12 (PMA) for optimal scheduling.

[0055] Available power and user-defined priorities are considered to schedule conflicting resource requests in a single clock cycle.

[0056] The memory device 4 allows up to four simultaneous operations, with a limit of one both for write and erase.

[0057] Figure 3 depicts the memory hierarchy and parallelism across the processing unit 1. The ports CP and DP are interfaced to the 64-bit, 800MB/s AHB system bus 6.

[0058] At a system clock rate of 100MHz each I/O port can independently operate at maximum speed. So, an aggregate peak read rate of 1.2GB/s can be sustained as it is limited by memory access time.

[0059] In the current implementation the e-FPGA reconfiguration takes 500μs at 100 MHz. 50MB/s average throughput out of the available 400MB/s are currently sustained by the e-FPGA configuration interface 7.

[0060] System performance is being evaluated for an image processing application (facial recognition) and a speech recognition application.

[0061] More than 20 specific instructions were designed as C/assembly-callable functions, automatically translated to RTL, then synthesized and mapped to the e-FPGA.

[0062] Figures 4 and 5 show two examples of specific microprocessor extensions.

[0063] Figure 4 relates to an eight-issue, eight-bit, L2 calculation accounts for 23 eight-bit arithmetic operations and six 64-bit operations requiring about 10k ASIC

equivalent gates.

[0064] Figures 5 relates to a datapath for an optimized fixed-point calculation of the square root accounts for twelve 32-bit operations for about 2k ASIC equivalent gates.

[0065] The overall performance improvements for the face recognition tasks are shown in the table of Figure 6.

[0066] Execution time is compared for 32-bit RISC with basic DSP extensions (MAC, zero-overhead loops, etc) and the same processor enhanced with application-specific instructions.

[0067] Measured speed-ups range from 1.8x to 10.6x (on the most-demanding task), with an overall improvement of 8.5x. It must be noticed that switching between algorithm stages requires only one reconfiguration of the e-FPGA. Reconfiguration time is negligible.

[0068] The speed-up factors take into account the possible multi-cycle clock penalty due to processor-FPGA synchronization in case of instruction extensions slower than the processor clock. Energy efficiency figures are reported in Figure 6 too.

[0069] As the average power consumption of the system extended with the e-FPGA is slightly higher (10-15%), the energy reduction for executing each of the tasks on its specific HW configuration (power-delay product improvement) results in an overall reduction of 6.7x.

[0070] Only one task showed slightly worse total execution energy, though showing benefits on execution speed.

[0071] Last column of Figure 6 reports the energy-delay improvement of each specific HW configuration compared to the general-purpose counterpart. Energy required for e-FPGA reconfiguration is always negligible.

[0072] Measurements show the best energy efficiency in the range of several MOPS/mW at 1.8V supply. It lies between conventional ASIP/DSP and dedicated configurable hardware implementations.

[0073] The full-processing unit on a single chip is implemented in a 0.18μm, 2PL-6ML CMOS embedded Flash technology, chip area is 70mm^2, technology and device characteristics are summarized in Figure 6 while a chip micrograph is shown in Figure 7.

Claims

1. A dynamically reconfigurable processing unit (1) including an embedded Flash memory device (3) for non-volatile storage of code, data and bit-streams, the unit (1) being integrated into a single chip together with a microprocessor (2) core, further comprising an S-RAM based embedded FPGA unit structured for FPGA reconfigurations having a specific programming interface (7) connected to a port (FA) of said Flash memory device (4) through a DMA channel (8).

2. A dynamically reconfigurable processing unit according to claim 1, wherein said DMA channel (8) handles the bitstream transfer while said microprocessor (2) fetches instructions and data from different Flash memory ports of said Flash memory device (4); a wide code port (CP) and a data port (DP).

3. A dynamically reconfigurable processing unit according to claim 2, wherein said Flash memory device (4) includes a modular array structure (13) comprising N memory blocks (16), and wherein a power block (10), including charge pumps, is shared among different flash memory modules (16) through a PMA arbiter (12) in a multi-bank fashion.

4. A dynamically reconfigurable processing unit according to claim 1, wherein said embedded FPGA unit (3) exploits the following functions:

   iv) extension of the processor datapath supporting a set of additional special-purpose C-callable microprocessor instructions;

   v) bus-mapped coprocessors, connected to the system bus through a master/ slave interface;

   vi) flexible I/O to connect external units or sensors with application-specific communication protocols.

5. A dynamically reconfigurable processing unit according to claim 2, wherein said Flash memory device (4) includes at least three different access ports, each for a specific function:

   - said code port (CP) optimized for random access time and the application system;
   - said data port (DP) allowing an easy way to access and modify application data; and,
   - said FPGA port (FP) offering a serial access for a fast download of bit streams for an embedded FPGA (e-FPGA) configurations.

6. A dynamically reconfigurable processing unit according to claim 2, wherein said third port (FP) comprises four configuration registers replicating the information stored in said code port (CP) that must be used in order to write e-FPGA configurations data.

7. A dynamically reconfigurable processing unit according to claim 5, wherein said third port (FP) uses a chip select to access in the addressable memory space and a burst enable to allow burst serial access.

8. A dynamically reconfigurable processing unit ac-

cording to claim 1, wherein said connection between said interface (7) and said port (FA) is provided by a local bus (6).

9. A dynamically reconfigurable processing unit according to claim 5, wherein said Flash memory device (4) includes four modules (16) each arranged in at least three programmable user-defined partitions, each one devoted to a corresponding port.

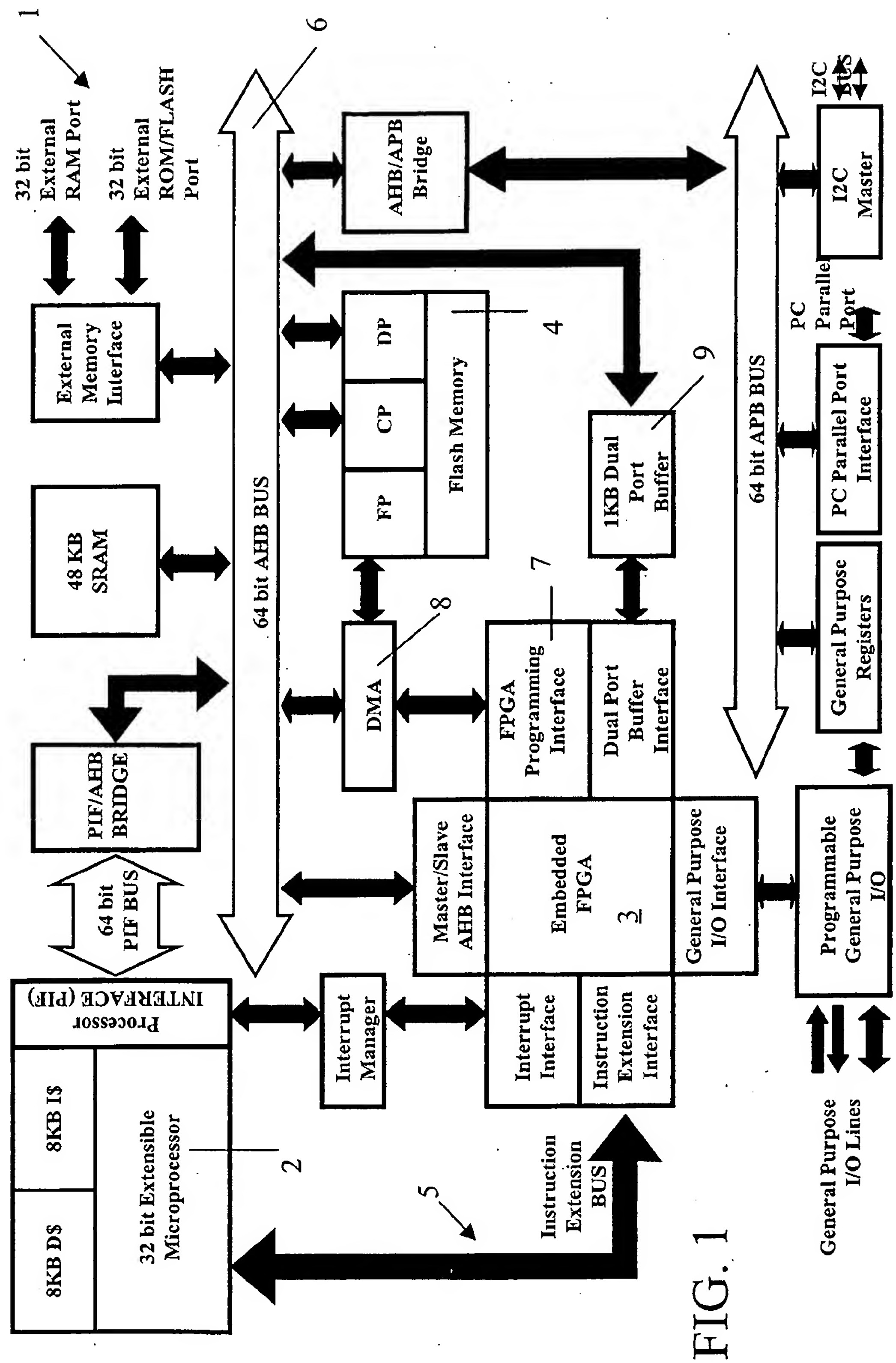

1
32 bit External RAM Port
32 bit External ROM/FLASH Port
External Memory Interface
48 KB SRAM
PIF/AHB BRIDGE
64 bit PIF BUS
Processor INTERFACE (PIF)
8KB I$
8KB D$
32 bit Extensible Microprocessor
2
64 bit AHB BUS
6
AHB/APB Bridge
DP
CP
FP
Flash Memory
4
DMA
8
Interrupt Manager
5
Instruction Extension BUS
Master/Slave AHB Interface
Embedded FPGA
3
FPGA Programming Interface
7
Dual Port Buffer Interface
1KB Dual Port Buffer
9
Interrupt Interface
Instruction Extension Interface
General Purpose I/O Interface
64 bit APB BUS
I2C BUS
I2C Master
PC Parallel Port
PC Parallel Port Interface
General Purpose Registers
Programmable General Purpose I/O
General Purpose I/O Lines

FIG. 1

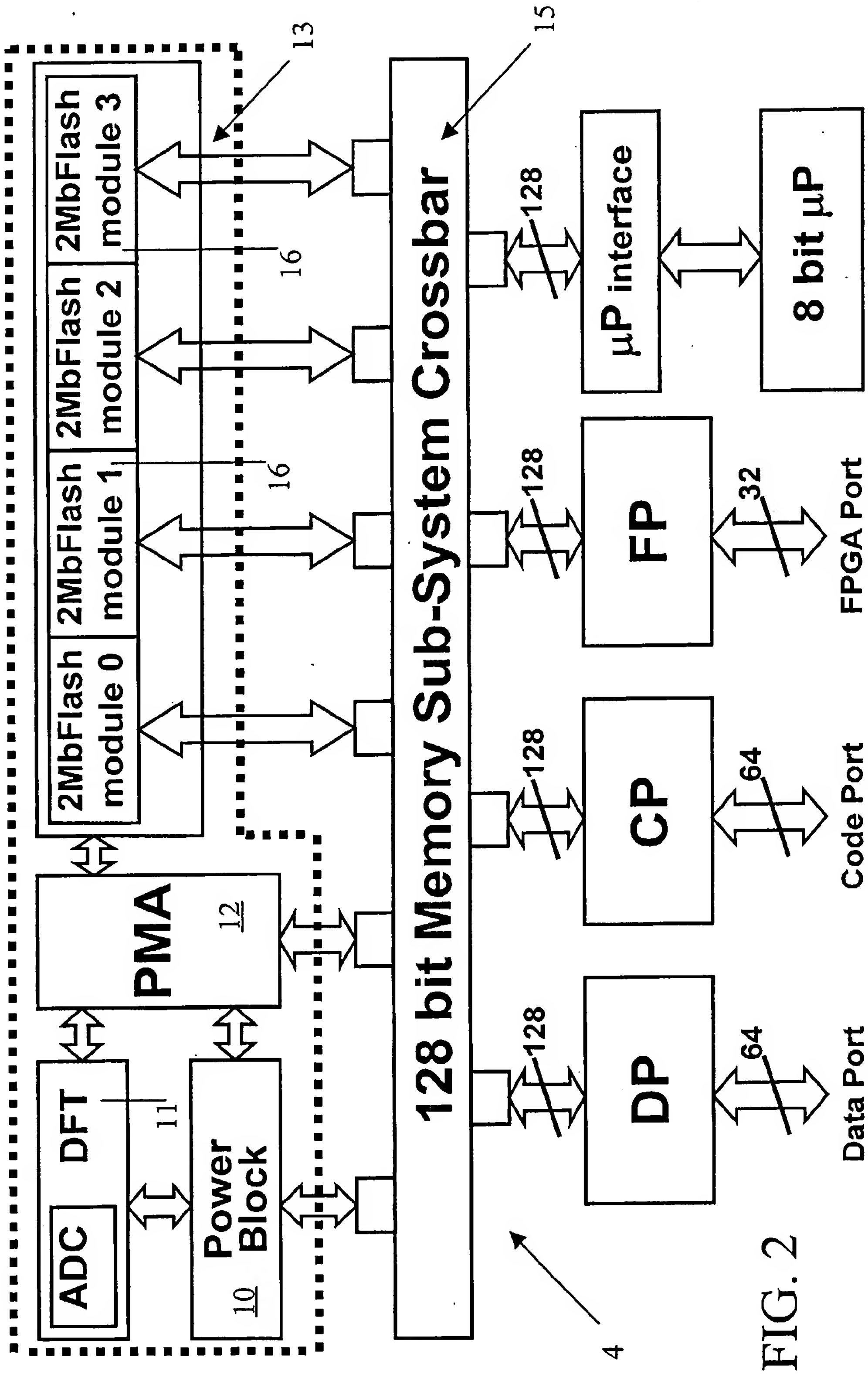

ADC
DFT
11
Power Block
10
PMA
12
2MbFlash module 0
2MbFlash module 1
2MbFlash module 2
2MbFlash module 3
16
16
13
15
128 bit Memory Sub-System Crossbar
128
128
128
128
DP
CP
FP
µP interface
8 bit µP
64
64
32
Data Port
Code Port
FPGA Port
4

FIG. 2

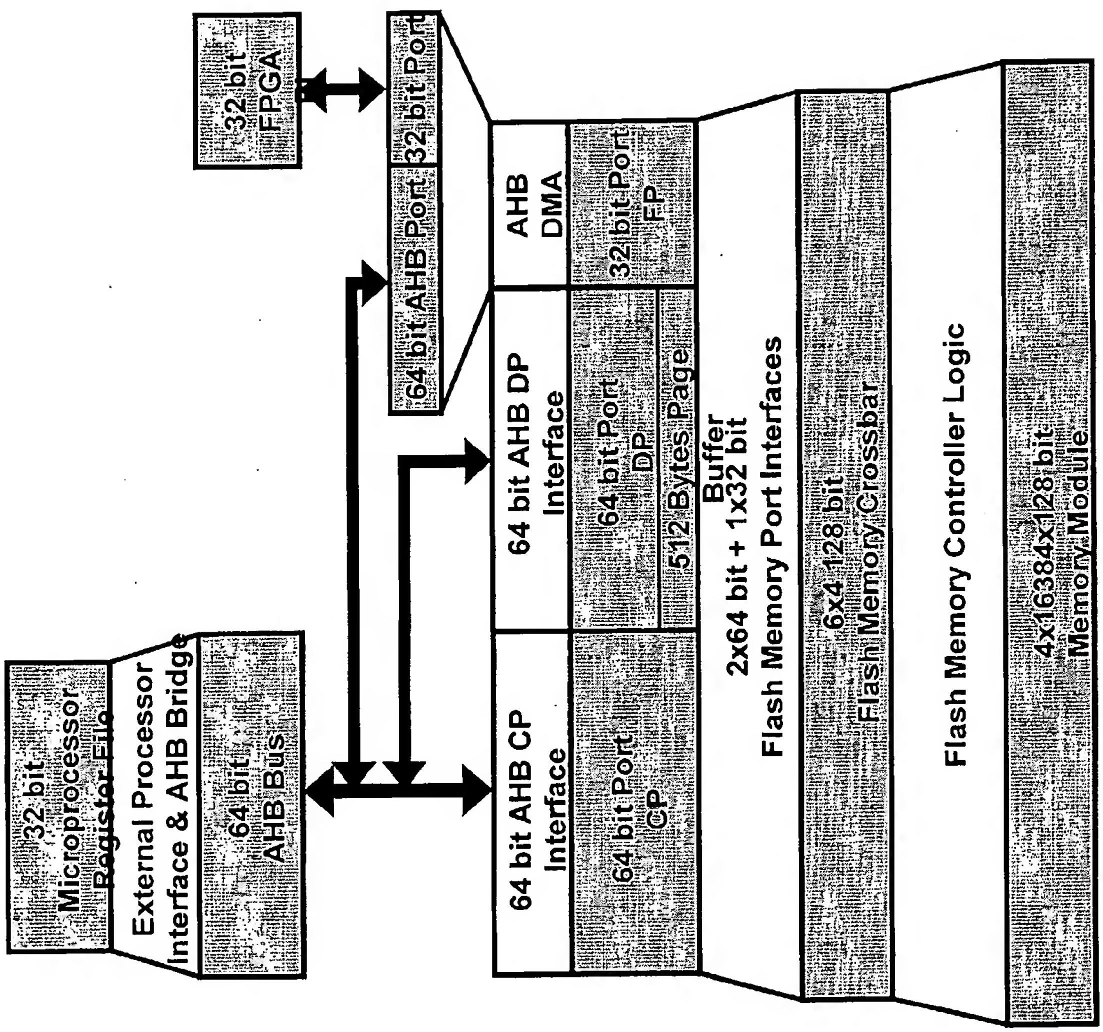
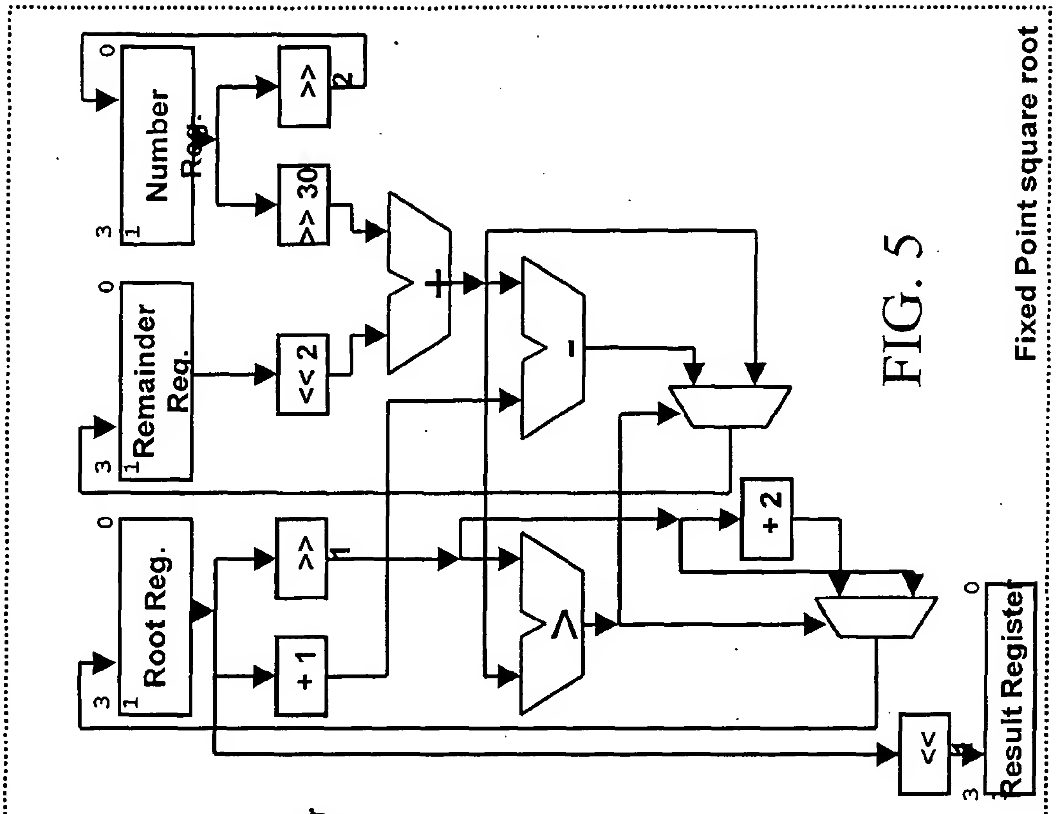
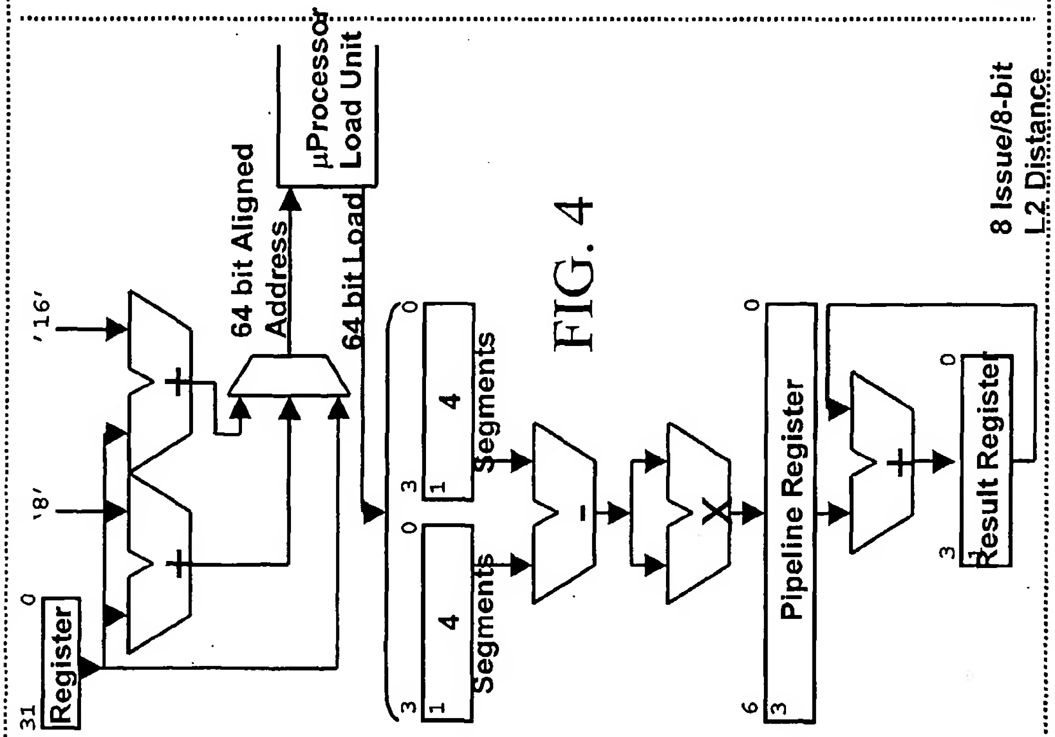

32 bit Microprocessor Register File
External Processor Interface & AHB Bridge
64 bit AHB Bus
32 bit FPGA
64 bit AHB Port
32 bit Port
64 bit AHB CP Interface
64 bit AHB DP Interface
AHB DMA
64 bit Port CP
64 bit Port DP
32 bit Port FP
512 Bytes Page Buffer
2x64 bit + 1x32 bit
Flash Memory Port Interfaces
6x4 128 bit Flash Memory Crossbar
Flash Memory Controller Logic
4x16384x128 bit Memory Module

FIG. 3



Root Reg.
Remainder Reg.
Number Reg.
>> 2
>> 30
<< 2
>> 1
+ 1
+ 2
<<
Result Register
Fixed Point square root

FIG. 5



Register
'8'
'16'
64 bit Aligned Address
µProcessor Load Unit
64 bit Load
4 Segments
4 Segments
Pipeline Register
Result Register
8 Issue/8-bit L2 Distance

FIG. 4

Algorithm Stage	RISC with basic DSP instruction support	RISC with Microprocessor Extensions	Speed-Up	Energy Gain	Energy Efficiency Gain (Energy x Delay)
Bayer filter	58 msec	24.7 msec	x 2.3	x 1.4	x 3.22
Edge detection	4.5 msec	2.5 msec	x 1.8	x 0.95	x 1.71
Face detection	1.5 sec	382 msec	x 4	x 2.9	x 11.6
Face recognition (Twenty face database)	9.15 sec	860 msec	x 10.6	x 9	x 95.4
Totals	**10.7 sec**	**1.26 sec**	**x 8.5**	**x 6.7**	

FIG. 6

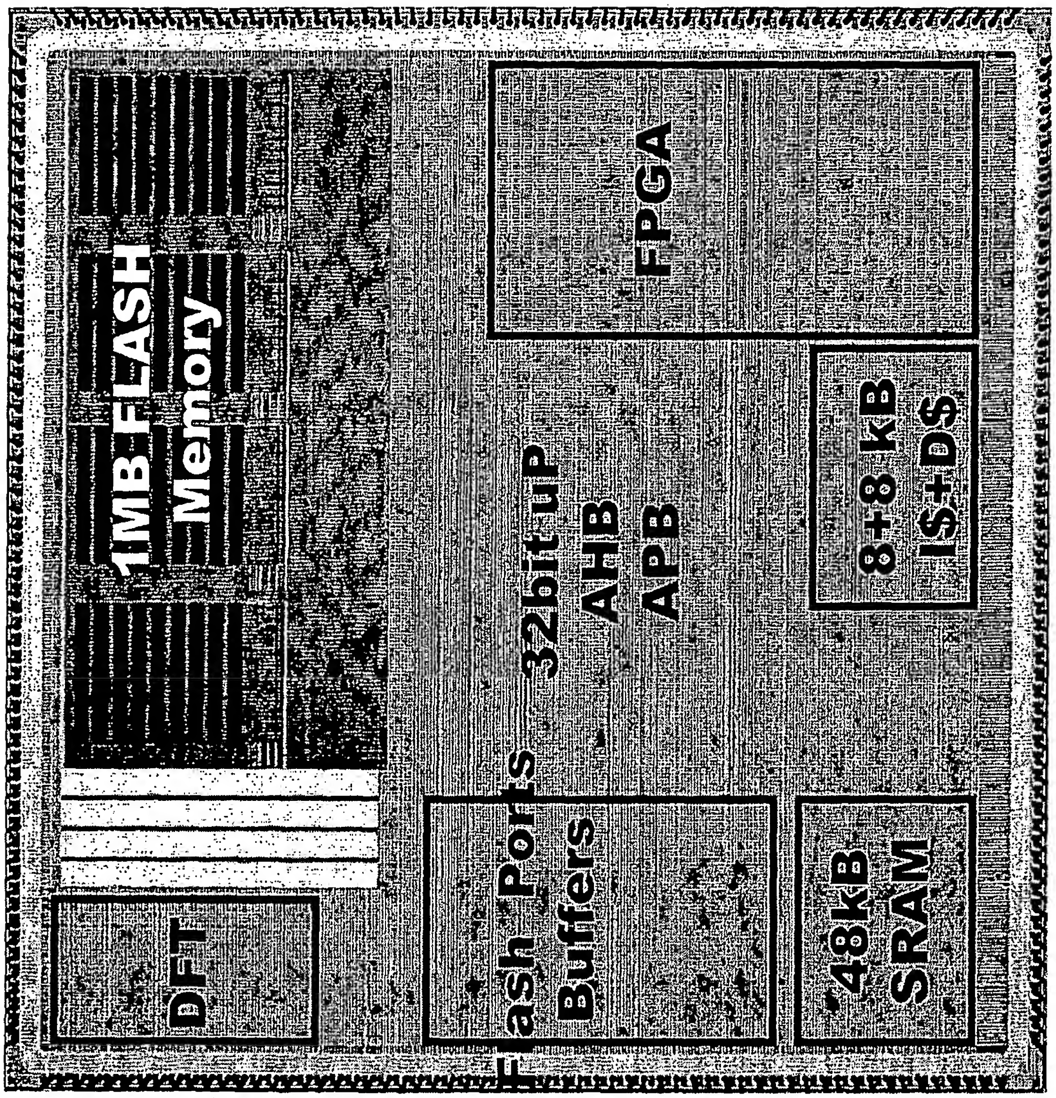

1MB FLASH
Memory
DFT
Flash Ports
Buffers
32bit uP
AHB
APB
FPGA
8+8 kB
I$+D$
48kB
SRAM

FIG. 7

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 03 42 5054

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
X	US 6 467 009 B1 (KRISHNAMURTHY SRIDHAR ET AL) 15 October 2002 (2002-10-15) * figures 1,2 * * column 34, line 66 - column 35, line 15 * * column 4, line 48 - line 50 *	1-9	G06F15/78 G11C16/30
A	DAVID R ET AL: "DART: a dynamically reconfigurable architecture dealing with future mobile telecommunications constraints" PARALLEL AND DISTRIBUTED PROCESSING SYMPOSIUM., PROCEEDINGS INTERNATIONAL, IPDPS 2002, ABSTRACTS AND CD-ROM FT. LAUDERDALE, FL, USA 15-19 APRIL 2002, LOS ALAMITOS, CA, USA,IEEE COMPUT. SOC, US, 15 April 2002 (2002-04-15), pages 156-163, XP010591206 ISBN: 0-7695-1573-8 * page 3, left-hand column, line 19 - line 22; figure 3 *	1,2	
A	US 5 693 570 A (LEE DOUGLAS J ET AL) 2 December 1997 (1997-12-02) * figure 2.A *	3	TECHNICAL FIELDS SEARCHED (Int.Cl.7) G06F G11C
A	"XC6200 Field Programmable Gate Arrays Product Description" , XILINX , 24-APRIL-1997 XP002262656 * page 31 *	7	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	26 November 2003	Bosch Vivancos, P

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 03 42 5054

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

26-11-2003

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 6467009	B1	15-10-2002	AU	6431999 A	01-05-2000
			GB	2361559 A ,B	24-10-2001
			WO	0022546 A2	20-04-2000
US 5693570	A	02-12-1997	US	5621685 A	15-04-1997
			US	5508971 A	16-04-1996
			US	5563825 A	08-10-1996
			US	5568424 A	22-10-1996
			US	5592420 A	07-01-1997

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82